AMENDMENTS TO THE SPECIFICATION

In the Specification:

Please replace the paragraph beginning at page 2, line 12 with the following amended paragraph:

The feature vectors can represent any number of available features extracted through known feature extraction methods such as Linear Predictive Coding (LPC), LPC-derived cepstrum, Perceptive Linear Prediction (PLP), auditory model, and Mel-Frequency Cepstrum Coefficients (MFCC).

Please replace the paragraph beginning at page 2, line 28 with the following amended paragraph:

The present invention provides for a system and method that facilitate modeling speech dynamics based upon a speech model, called the segmental switching state space model, that employs model parameters that characterize some aspects of the human speech articulation process. These model parameters are modified based, at least in part, upon a variational learning technique.

Please replace the paragraph beginning at page 3, line 3 with the following amended paragraph:

In accordance with an aspect of the present invention, novel and powerful variational expectation maximization (EM) algorithm(s) for the segmental switching state space models used in speech applications, which are capable of capturing key internal (or hidden) dynamics of natural speech production, are provided. Hidden dynamic models (HDMs) have recently become a class of promising acoustic models to incorporate crucial speech-specific knowledge and overcome many inherent weaknesses of traditional HMMs. However, the lack of powerful and efficient statistical learning

algorithms is one of the main obstacles preventing them from being well studied and widely used. Since exact inference and learning are intractable, a variational approach is taken to develop effective approximate algorithms. The present invention implements the segmental constraint crucial for modeling speech dynamics and provides algorithms for recovering hidden speech dynamics and discrete speech units from acoustic data only. Further, the effectiveness of the algorithms developed is verified by experiments on simulation and Switchboard speech data.

Please replace the paragraph beginning at page 5, line 8 with the following amended paragraph:

The system 100 can utilize powerful variational expectation maximization (EM) algorithm(s) for the segmental switching state space models used in speech applications, which are capable of capturing key internal (or hidden) dynamics of natural speech production. The system 100 overcomes inherent weakness of traditional HMMs by employing efficient statistical learning algorithm(s). Since exact inference and learning are intractable, in accordance with an aspect of the present invention, the system 100 utilizes a variational approach ~~is taken~~ to develop effective approximate algorithms. Thus, the system can implement the segmental constraint crucial for modeling speech dynamics and provides algorithms for recovering hidden speech dynamics and discrete speech units from acoustic data only.

Please replace the paragraph beginning at page 5, line 18 with the following amended paragraph:

The system 100 includes an input component 110 that receives acoustic data. For example, the input component 110 can convert an analog speech signal into a series of digital values. The system further includes a model component 120 that models speech. The model component 120 receives the acoustic data from the input component 110. The model component 120 then recovers speech from the acoustic data based, at least in part, upon a model having model parameters including the parameters which characterize

aspects of the unobserved dynamics in speech articulation and the parameters which characterize the mapping relationship from the unobserved dynamic variables to the observed speech acoustics. The model parameters are modified based, at least in part, upon a variational learning technique as discussed below.

Please replace the paragraph beginning at page 5, line 28 with the following amended paragraph:

In one example, the model component 120 employs an HDM in a form of switching state-space models for speech applications. The state equation and observation equation are defined to be:

$$\mathrm{x}_n = \mathrm{A}_s \mathrm{x}_{n-1} + (\mathrm{I} - \mathrm{A}_s)\mathrm{u}_s + \mathrm{w}, \qquad (1)$$

$$\mathrm{y}_n = \mathrm{C}_s \mathrm{x}_n + \mathrm{c}_s + \mathrm{v}, \qquad (2)$$

where *n* and *s* are frame number and phone index respectively, *x* is the hidden dynamics and *y* is the acoustic feature vector (such as MFCC). For example, the hidden dynamics can be chosen to be the articulatory variables, or to be the variables for the vocal-tract-resonances (VTRs) which are closely related to the smooth and target-oriented movement of the articulators. The state equation (1) is a linear dynamic equation with phone dependent system matrix $\mathbf{A}_s$ and target vector $\mathbf{u}_s$ and with built-in [[build-in]] continuity constraint across the phone boundaries. The observation equation (2) represents a phone-dependent VTR-to-acoustic linear mapping. The choice of linear mapping is mainly due to the difficulty of algorithm development. The resulting algorithm can also be generalized to mixtures of linear mapping and piece-wise linear mapping within a phone. Further, Gaussian white noises **w** and **v** can be added to both the state and observation equations to make the model probabilistic. **C** and c represent the parameters responsible for the mapping from the VTRs to the acoustic feature vector.

Please replace the paragraph beginning at page 7, line 21 with the following amended paragraph:

The idea is to choose the approximate posterior q to approximate the true posterior $p(s_{1:N}, \boldsymbol{x}_{1:N} \mid \boldsymbol{y}_{1:N}))$ with a sensible and tractable structure and optimize it by minimizing its Kullback-Liebler (KL) distance to the exact posterior. It turns out that this optimization can be performed efficiently without having to compute the exact (but intractable) posterior.

Please replace the paragraph beginning at page 8, line 2 with the following amended paragraph:

As discussed previously, in one example, the system 100 employs an approximation based, at least in part, upon a mixture of Gaussian (MOG) posterior. Under this approximation q is restricted to be:

$$q(s_{1:N}, \mathrm{x}_{1:N}) = \prod_n q(\mathrm{x}_n \mid s_n) q(s_n), \tag{5}$$

For purposes of brevity, the dependence of the q's on the observation **y** is omitted but always implied.

Please replace the paragraph beginning at page 8, line 10 with the following amended paragraph:

Minimizing the KL divergence between q and p is equivalent to maximizing the following function F,

$$\begin{aligned} F[q] = \sum_{s_{1:N}} \int d\mathrm{x}_{1:N} q(s_{1:N}, \mathrm{x}_{1:N}) \cdot \\ [\log p(\mathrm{y}_{1:N}, \mathrm{x}_{1:N}, s_{1:N}) - \log q(s_{1:N}, \mathrm{x}_{1:N})], \end{aligned} \tag{6}$$

which is also a lower bound of the likelihood function and will be subsequently used as the objective function in the learning (M) step.

Please replace the paragraph beginning at page 13, line 21 with the following amended paragraph:

1. PARAMETER INITIALIZATION

Please replace the paragraph beginning at page 14, line 8 with the following amended paragraph:

2. SEGMENTAL CONSTRAINT

Please replace the paragraph beginning at page 17, line 22 with the following amended paragraph:

At 830, an approximation of a posterior distribution based upon a mixture of Gaussian posteriors is calculated. For example, calculation of the approximation of the posterior distribution can be based, at least in part, upon Equation (5). At 840, the model parameter(s) are modified based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Leibler distance of the approximation from an exact posterior distribution.

Please replace the paragraph beginning at page 18, line 3 with the following amended paragraph:

At 930, an approximation of a posterior distribution based upon a mixture of hidden Markov model posteriors is calculated. For example, calculation of the approximation of the posterior distribution can be based, at least in part, upon Equation (20). At 940, the model parameter(s) are modified based, at least in part, upon the

calculated approximated posterior distribution and minimization of a Kullback-Leibler distance of the approximation from an exact posterior distribution.